

# High dynamic range imaging with a single exposure-multiplexed image using smooth contour prior

Mushfiqur Rouf; Department of Computer Science, University of British Columbia; Vancouver, BC

Rabab K. Ward; Department of Electrical and Computer Engineering, University of British Columbia; Vancouver, BC

## Abstract

*Photography is an ever-advancing technology: pixels are getting more densely packed, colors are going deeper, and light sensitivity is going higher. But the camera standard dynamic range largely remains less than the full brightness range of a natural scene. In the days of film cameras people have used the multi-exposure high dynamic range imaging to capture the full dynamic range although only limited for static scenes. However, with digital photography, the exposure index can be spatially multiplexed on a sensor, which adapts the essence of the multi-exposure technique for single-image high dynamic range imaging. A sophisticated postprocessing is required; and existing techniques either lose final image resolution or take a long time to run. In this paper, we propose a novel method for this postprocessing: we use a robust global optimization method and a fast edge-preserving interpolation technique for full resolution exposure-multiplexed high dynamic range imaging.<sup>1</sup>*

**Keywords:** High dynamic range, demosaicking, edge-directed.

## Introduction

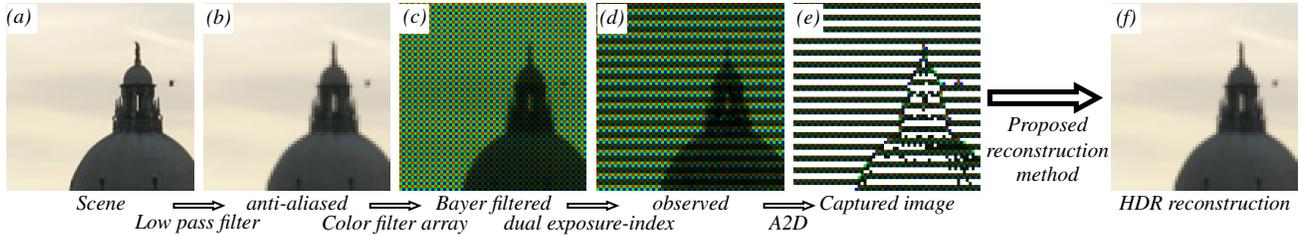
Camera sensors can faithfully capture only a limited range of brightness variation; this range is known as the dynamic range (DR) of a camera. Each pixel of the sensor acts as a photon counter; a bright scene point will send more photons than a pixel can count effectively saturating the sensor. Measurements from all pixels of the sensor is fed into an analog-to-digital converter which applies an analog gain or exposure index (or film speed, or more commonly referred to as the “ISO” [1]). The absolute range of values faithfully captured in an exposure depends on the shutter speed, the aperture size and the exposure index. Note that the relative brightness ratio between the brightest and the darkest measured pixel values remains constant, and this constant is the DR: the ratio between the clipping “ceiling” and the noise “floor”.

High dynamic range imaging (HDRI) techniques use these standard dynamic range (SDR) cameras and capture high dynamic range images. The most well-known high dynamic range imaging technique involves capturing multiple images of the same scene, each image captured with a different camera sensor exposure settings – shutter speed [2, 3], aperture [4] or exposure index [5]. This way bright areas will be captured by an image captured with a short exposure setting while the dark areas will be captured by another image of the sequence captured with a long exposure setting. When all this captured standard dynamic range image data is combined offline, a high dynamic range image is produced. This method is guaranteed to produce the best high dynamic range im-

ages; however, there are strict limitations that have to be met for the sake of high quality: the camera cannot move, the scene has to be static, and the lighting conditions cannot vary between exposures. If one or more of these constraints get violated, ghosting will result when merging the image sequence into a HDR image. A recent method [5] aims at capturing multiple very short exposures in rapid succession aka “burst mode” in order to reduce the misalignment problem but it fails to capture details in the dark regions effectively. Deghosting HDR images has been extensively explored [6], however it still remains largely an open problem. To avoid ghosting, one can use multiple cameras sharing the same optical axis and imaging with different exposure settings to capture the exposure sequence at once. However, this is an expensive solution, this introduces additional optical components on the light path which adds glare to the system, and this setup requires perfect calibrations; slightest relative camera motions can result in a misalignment which cannot be precalibrated.

Some recent image sensors as a middle path have introduced an optional feature that allows a user to perform a resolution-DR tradeoff; in exchange for a potential resolution loss a user can choose to capture a wider DR in a single capture, which can be particularly useful in situations where taking multiple exposures is not an option. These sensors can spatially multiplex exposure index. An exposure-multiplexed image contains more scene information compared to a uniform-exposure index capture. Pixels that are exposed at same exposure can observe the same brightness variation, while other pixels with a different exposure setting can observe a different range of brightness values of the same scene. Effectively, this process allows capturing “multiple” exposures on a single exposure on a single sensor, each of which is taken with a different exposure setting. This however, comes at the cost of resolution loss. The general idea is: in the same single exposure of an imaging sensor, different groups of pixels use different exposure indices. While each group individually still has the original standard dynamic range (SDR), when the data is demultiplexed the combined image effectively has a wider DR. First such method [7] proposed to lay spatially varying neutral-density (ND) filters for exposure index multiplexing. However, this is not a desirable solution since (1) once installed it cannot be changed or adapted for scene types and (2) ND filter will absorb a large portion of the light incident on the sensor. Some recent sensors can *multiplex* electronic shutter speed [8] or exposure index [9] between row-pairs, i.e., rows 1, 2, 5, 6, 9, ... can be set up with one setting, and rows 3, 4, 7, 8, 11, ... with the other. Varying shutter speed has the advantage of having uniform noise properties throughout the sensor for any one exposure, but this also poses the risk of ghosting [8] as the slower shutter speed will result in a larger blur in case an object moves in the scene. On

<sup>1</sup>This work is supported in part by the Institute for Computing, Information and Cognitive Systems (ICICS) at UBC.



**Figure 1:** The image formation model. Light from the scene gets focused onto the image sensor for acquisition. This optical data goes through a number of transformations before converting into digital sensor data. The scene (a) is optically band-limited (b) for avoiding aliasing using a spatial low-pass filter. Each sensor pixel can only measure in either red, or green or blue (c), as set out by the color filter array pasted on top of the image sensor. This band-limited and color-filtered optical information then gets captured by the sensor into an array of noisy analog measurements, which is then enhanced using per-pixel exposure index and converted to digital image data by an A2D converter. Pairs of rows alternate between a low exposure index and a high exposure index (d). The low sensitivity pixels can capture the bright scene areas well but the dark and shaded scene areas lose detail due to noise. The high sensitivity pixels can capture the relatively darker scene areas better, but get clipped due to saturation in bright parts of the image (e). Finally, we solve an inverse problem and obtain a maximum-a-posteriori (MAP) estimate of the unknown band-limited latent image (b).

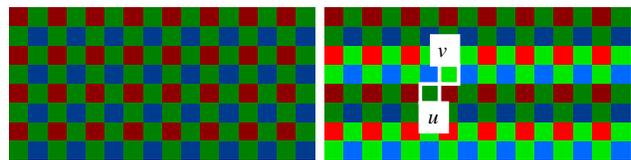
the other hand, modifying exposure index can vary sensor noise properties but no temporal alignment is required since both sets of pixels come with the same exposure begin and end times.

In this paper, we aim to recover the full resolution image from such an exposure-multiplexed raw image. In this image reconstruction problem, depending on the choice of exposure indices used in the exposure-multiplexed imaging, up to half of the raw pixels might have no image data (Fig. 1). The reconstruction of this missing data is an underdetermined inverse problem. This problem has similarities to other inverse imaging problems such as single-image super-resolution, demosaicking and inpainting, in that the input image has missing pixels, but none of these methods would directly fit our exposure-multiplexed imaging problem. For exposure-multiplexed imaging, several methods have looked into recovering the lost resolution: Magic Lantern an open source community has developed tools to take advantage of the exposure-multiplexed imaging capability on certain Canon cameras [10], but they use a simple edge-guided interpolation only. Other methods apply local statistical properties for spatial interpolation [11]. Instead, in this paper, we observe the similarities between our problem and the single-image super-resolution and we adapt [12] for our problem.

## Exposure-multiplexed HDR imaging

For this work, we use exposure-multiplexed imaging [9] in which the sensor can acquire more information on a single image compared to a single image obtained with conventional imaging with a uniform exposure index. Intuitively, a direct impact of the dual exposure-index imaging is the loss of resolution: none of the pixel sets contains the full image data for any given sensitivity setting; instead complementary sets of pixels are observing possibly complementary ranges of intensity values. We observe that this problem is similar to a superresolution problem: the forward dual exposure-index image formation process is similar to performing a spatial subsampling of the HDR image for each intensity range. This observation motivated us this research; in particular, we base our work on recent single-image super-resolution research [12].

We derive our proposed method in the following discussion. First we describe our optical encoding step in which we capture a single SDR image  $g$ . The image formation model from this step then leads to the computational decoding step in which we solve a convex optimization problem to obtain the full resolution HDR



(a) Conventional Bayer pattern (b) exposure-index multiplexed

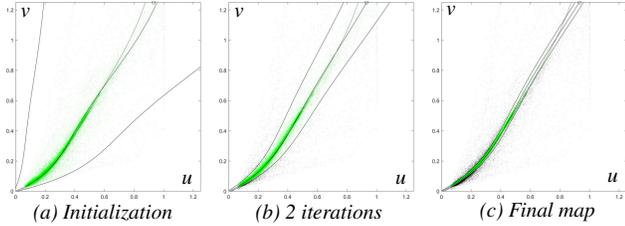
**Figure 2:** Color imaging. The photosensitive component of sensor pixels are fabricated with a uniform wavelength sensitivity profile. For color imaging, a color filter array is placed on top of a sensor. **Left:** most common color filter pattern is known as the Bayer pattern. Bayer pattern repeats a  $2 \times 2$  “RGGB” pattern over the entire sensor. **Right:** dual exposure-index imaging modifies the sensitivity. We show the high exposure index with brighter colors. Since the Bayer pattern has a basic block size of  $2 \times 2$ , exposure index is alternated only every two rows. The pixels  $u$  and  $v$  denotes two neighboring pixels with the same color filter but opposite exposure indices. Since natural images are mostly flat, it is likely that both of these pixels receive the same incident light intensity. We use all such  $\{u, v\}$  pairs to compute a polynomial estimate of the intensity-response curve  $R$  using a robust polynomial fitting.

image  $f$  from the captured SDR image  $g$ . This inverse problem is underdetermined, and so we obtain a maximum-a-posteriori (MAP) estimator of the unknown HDR image of interest  $f$ . Since MAP estimation leverages the Bayesian argument, it is crucial to have appropriate prior models of  $f$ ; we use a combination of two natural image priors: the first one is the sparse gradients prior [13, 14] which models the edge sparsity in natural images, and the second one is a novel prior we propose in this paper based on our prior work on single-image super-resolution [12].

## Forward model: image formation

The forward model describes the optical pipeline. We assume that the underlying unknown HDR image  $f$  has three color channels: RGB, and  $n$  pixels. The camera observes a SDR image we denote by  $g$ . Since light is linear, a system of linear equations can present this transformation. For the ease of derivation, we vectorize these images, i.e.,  $f \in \mathbb{R}^{3n}$  and  $g \in \mathbb{R}^n$  where  $n$  is the number of pixels in the captured image. We model the forward image formation process as a linear map from  $3n$  real numbers, through the optical processes, finally transforming into the  $n$  numbers of  $g$ .

In exposure multiplexing, the conventional imaging process remains mostly unchanged. The only modification is how sensor exposure index is set (Fig 1). We briefly describe this image



**Figure 3:** Preprocessing: corrections for the nonlinear response curve using RANSAC polynomial fitting. We initialize it the polynomial fitting using all points. This means that a wide dispersion is allowed initially, admitting the outliers which we subsequently reject iteratively. In each iteration, we reduce the allowed dispersion, find the accepted points that are within the allowed dispersion, and refit the polynomial with only these accepted points (b). We iterate until at least 25% points are rejected (c).

formation model to introduce the notations.

**Antialiasing filtering.** filter is necessary to remove aliasing artifacts in the captured image. Most image sensors implement this filtering by optically diffusing the incident image just before it hits the sensor. Effectively, it is a small Gaussian blur  $\rho$ .

**Color filtering.** A color sensor has a red-green-blue color filter array pasted on top. As a result, each pixel can only observe one of the three colors. The most common pattern of these three colors is called Bayer pattern [15] (Fig. 2a). RGB image on to the single-channel Bayer image as  $M : 3n \rightarrow n$  which is a mostly sparse  $3n \times n$  matrix with only one 1 per row—a color channel selector.

**Noise.** Multiple physical and electronic processes cause the noise in image acquisition. For SDR cameras, we can assume that the noise is additive white Gaussian noise (AWGN), i.e., is distributed as a zero-mean Gaussian with variance  $\sigma^2$ :  $\eta \sim \mathcal{N}(0, \sigma^2)$ .

**Intensity-response.** We denote the mapping from input intensity to output readout level by  $R : \mathbb{R}^+ \rightarrow [0, 1]$ . This response is almost linear. It only differs from the ideal linear response on either end of the observable intensity range: close to 0 and 1.

**Exposure index.** Exposure index  $\gamma$  denotes by what factor the analog electronic signal from a pixel is boosted before converting it to digital. Exposure multiplexing needs to take into account the basic block structure of the color filter array (Fig. 2b).

**Sensor saturation.** Sensor pixel electronics have a physical capacity that limits the maximum brightness a sensor pixel can observe. Beyond this level the brightness signal gets clipped at that maximum. W.l.o.g., we set this highest value to 1. Then sensor clipping is simply an operator  $\min(1, \cdot)$ .

The forward image formation process is therefore,

$$g = \min(1, R(\gamma M(\rho \otimes f) + \eta)) \quad (1)$$

where  $\otimes$  denotes convolution. Since saturated pixels provide no information, we ignore these rows, and rearrange (1) to get,

$$g_1 \equiv \frac{1}{\gamma} R^{-1}(g) = M(\rho \otimes f) + \eta. \quad [g < 1] \quad (2)$$

Most terms are known in this equation:  $g$  is the known captured image, per-pixel sensitivity  $\gamma$  is preset by the user before capturing the photo, the color map  $M$  is a fixed color filter array which is known,  $\rho$  is fixed for every sensor type and can be measured. Also, we accurately estimate the response curve  $R$  from the single image as we describe below. The only unknown quantities are the unknown HDR image  $f$  we solve for and the noise term  $\eta$ .



(a) Before correction (b) After correction  
**Figure 4:** Preprocessing: comparison of an unprocessed image assuming linear intensity response (a) with the preprocessed image (b). The insets clearly show that because of exposure index multiplexing high exposure index and low exposure index areas next to each other do not match. The mismatch shows up as fringes.

### Correction for exposure-index multiplexing

It is possible to estimate  $R$  by calibrating the camera response curve separately for various exposure indices. However, noise and other electronic factors make it difficult to use a single response curve in an artifact-free manner. Instead, we use a RANSAC-like iterative polynomial-fitting to calibrate for the intensity mapping  $R$  from the unclipped pixels of the single input image provided to our algorithm. We assume the intensity mapping  $R$  is a monotonic zero-crossing polynomial,

We perform this calibration once per channel. For each channel, we gather all pairs of neighboring pixel with different exposure index and no clipping. Let, for one pair of nearby pixels with the same color, the original anti-aliased intensities are  $u$  and  $v$  (Fig. 2b), and let the exposure indices are 1 and  $\gamma > 1$  respectively. We assume that  $u$  does not change much since the response curve is mostly linear. Then according to the forward model, the observed values are  $R(u) \approx u$  and  $R(\gamma v)$  ignoring noise and clipping. However, because the response curve  $R$  is not linear,

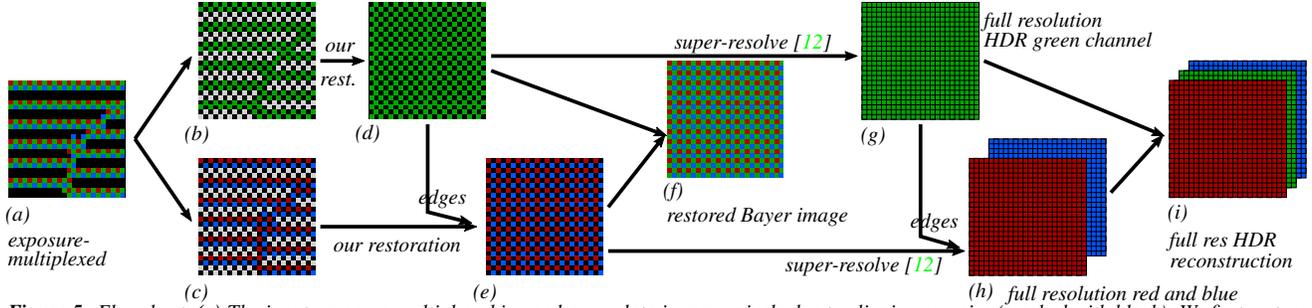
$$R(\gamma v) / \gamma \neq (u). \quad (3)$$

We know from the sparse gradients prior [13, 14] that natural images for the most part is flat, except at sparse sharp edges in the scene. Therefore, except for a few outliers,

$$u \approx v. \quad (4)$$

We then seek the a fifth order polynomial to model  $R$  to solve (3) and (4). We perform a robust RANSAC-style polynomial fitting for obtaining the estimated curve. A least square fit only models Gaussian-distributed quantities (and is greatly affected by outliers). We initialize the curve by fitting all points. After this initialization, we reject more and more outliers in each iteration: we reduce sigma and recompute the curve until at least a large enough part (25% in experiments) of the pairs are rejected. This ensures a fit that is robust against outliers, i.e., edges. We denote the scaled captured image as  $g_1$ . This simplifies the forward model in (2),

$$g_1 = M(\rho \otimes f) + \eta, \quad [g < 1] \quad (5)$$



**Figure 5:** Flowchart. (a) The input exposure-multiplexed image has no data in many pixels due to clipping or noise (marked with black). We first restore the missing pixel values in the green channel (b) using our algorithm and obtain all green pixel values (d). We use the edge structures from this restored green channel to calculate smooth contour prior’s local anisotropic filters and use those to reconstruct the red and blue channels (c). The resulting image (e) contains all red and blue pixel information on the Bayer image. (d) and (e) combined constitutes the fully restored Bayer image (f). Going from this image to the full resolution image is essentially a demosaicking problem and we use [12] for this purpose. As in the restoration step, we first obtain the full resolution green channel (g), then we use the edge structure in this channel to guide the reconstruction of the other two channels (h). This gives us the reconstructed full resolution high dynamic range image (i).

### HDR reconstruction

The crux of our high dynamic range imaging method is reconstructing the information that is missing due to over-saturation (i.e., clipping) or under-saturation (i.e., noise) of image sensors (Fig. 1f). To obtain the full-resolution HDR image, we need to perform edge-preserving anisotropic filtering of the captured data.

As shown in Fig. 5, we first restore the green channel because it is more densely sampled by the Bayer filter, it is less hard to restore using single-channel data. We then use the restored green channel to guide the edge-preserving interpolation in the other two channels using [12] (Fig. 5). Below, we first formulate HDR reconstruction as a global optimization problem and then give our algorithm. In the next section we derive our smooth contour prior.

The image formation model (5) gives the forward model: the optical encoding. In order to estimate the unknown latent HDR image  $f$ , we perform a maximum-a-posteriori (MAP) estimation in the same fashion as the single image superresolution work [12]: we minimize noise  $\eta \sim \mathcal{N}(0, \sigma^2)$ , (i.e., underdetermined data-fitting) such that the following image priors (for well-posedness) are satisfied: (1) natural image gradients are sparse (the “sparse gradients prior” [16]), and (2) natural image edges have smooth contours (i.e., the “smooth contours image prior” [12]). From (5) we derive the minimization problem,

$$\min_f \underbrace{\frac{1}{2\sigma^2} \|g_1 - M(\rho \otimes f)\|^2}_{\text{data-fitting, } g < 1} + \underbrace{\lambda_{\text{TV}} \|\nabla f\|_{\text{TV}}}_{\text{sparse gradient}} + \underbrace{\lambda_{\mathcal{E}} \|f - \mathcal{E}(f)\|^2}_{\text{smooth contour}} \quad (6)$$

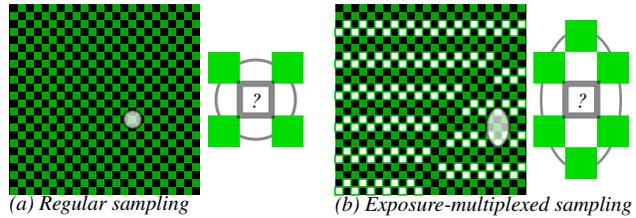
where  $\|\cdot\|_{\text{TV}}$  is the total variation norm,  $\mathcal{E}$  is the smooth-contour interpolation operator, and  $\lambda_{\text{TV}}$  and  $\lambda_{\mathcal{E}}$  are proportionality constants. The  $\mathcal{E}$  operator from [12] unmodified is not suitable for our problem; later in this text we derive a modification.

Although (6) is convex, solving it directly is not easy because of the mixed  $\ell_1$  and  $\ell_2$  norms. Instead, we solve the equivalent saddle point forms. First, let  $\mathbf{G}(\cdot)$  denote our data-fitting term,

$$\mathbf{G}(f) \equiv \frac{1}{2\sigma^2} \|g_1 - M(\rho \otimes f)\|^2 \quad (7)$$

$K_{\text{TV}}$  and  $K_{\mathcal{E}}$  are linear transformations from the primal (image) domain to respective dual domains defined as

$$K_{\text{TV}} f \equiv \lambda_{\text{TV}} \nabla f \quad (8)$$



**Figure 6:** Choice of support for the smooth contour prior. Only the green channel is shown. Extending this idea for other channels is trivial. (a) Green channel pixels are organized on a diagonal grid, shown in green. In conventional imaging, the sensor has a uniform exposure setting, and the green channel gets sampled evenly both horizontally and vertically. The support—the size of the local anisotropic filter kernel we estimate—uses a disk shape support on a grid turns into a 4-neighborhood on a grid. (b) In exposure-multiplexed imaging, the vertical sampling rate can be as low as half the horizontal sampling rate; the white pixels denote missing data. The anisotropic interpolation kernel we seek in this case needs to be twice as long in the vertical direction, to account for the half sampling rate. This way we are able to gather as much structural information as the regular sampling case. Taking into account green pixel locations, this amounts to the 6-neighborhood on the right.

$$K_{\mathcal{E}} f \equiv f - \mathcal{E}(f), \quad (9)$$

and  $\mathbf{F}_{\text{TV}}$  and  $\mathbf{F}_{\mathcal{E}}$  are functions defined as

$$\mathbf{F}_{\text{TV}} \equiv \|\cdot\|_{\text{TV}} \quad (10)$$

$$\mathbf{F}_{\mathcal{E}} \equiv \frac{\lambda_{\mathcal{E}}}{2} \|\cdot\|^2. \quad (11)$$

Now, rewriting the convex optimization problem (6), we get,

$$\arg \min_f \mathbf{G}(f) + \mathbf{F}_{\text{TV}}(K_{\text{TV}} f) + \mathbf{F}_{\mathcal{E}}(K_{\mathcal{E}} f), \quad (12)$$

which is equivalent to solving the primal-dual form,

$$\arg \min_f \mathbf{G}(f) + \arg \max_{y_{\text{TV}}} \langle K_{\text{TV}} f, y_{\text{TV}} \rangle - \mathbf{F}_{\text{TV}}^*(y_{\text{TV}}) + \arg \max_{y_{\mathcal{E}}} \langle K_{\mathcal{E}} f, y_{\mathcal{E}} \rangle - \mathbf{F}_{\mathcal{E}}^*(y_{\mathcal{E}}), \quad (13)$$

where  $\cdot^*$  denotes the convex conjugate of a function, and  $y_{\text{TV}}$  and  $y_{\mathcal{E}}$  are slack variables defined over respective dual domains. The algorithm we propose in this paper directly follows from [12]. We use the primal-dual algorithm [17] to solve the saddle-point formulation (13) of our original inverse problem (6).

**Algorithm 1** Exposure-multiplexed high dynamic range imaging using the smooth contour prior

**Require:**  $\sigma > 0$ ,  $\tau > 0$  and  $\theta$ .  $\bar{f}^{(0)} = f^{(0)} = \mathcal{E}(g_1), y_{\text{TV}}^{(0)} = 0, y_{\mathcal{E}}^{(0)} = 0$ .

- 1: **repeat**
- 2:  $y_{\text{TV}}^{(k+1)} \leftarrow \text{prox}_{\mathbf{F}_{\text{TV}}^*}^{\sigma} \left( y_{\text{TV}}^{(k)} + \sigma K_{\text{TV}} \bar{f}^{(k)} \right)$
- 3:  $y_{\mathcal{E}}^{(k+1)} \leftarrow \text{prox}_{\mathbf{F}_{\mathcal{E}}^*}^{\sigma} \left( y_{\mathcal{E}}^{(k)} + \sigma K_{\mathcal{E}} \bar{f}^{(k)} \right)$
- 4:  $f^{(k+1)} \leftarrow \text{prox}_{\mathbf{G}}^{\tau} \left( f^{(k)} - \tau \left( K_{\text{TV}}^T y_{\text{TV}}^{(k+1)} + K_{\mathcal{E}}^T y_{\mathcal{E}}^{(k+1)} \right) \right)$
- 5:  $\bar{f}^{(k+1)} \leftarrow f^{(k+1)} + \theta (f^{(k+1)} - f^{(k)})$ .
- 6: **until** convergence

This primal-dual optimization Algorithm 1 starts with some initial values for the unknown image  $f$  and the dual-domain slack variables  $y$ . The algorithm then proceeds with a series of generalized projections alternating between the primal and dual domains. These generalized projections are called “proximity” operators and are denoted by **prox**. Please consult [17] for detailed derivation of the method and convergence guarantees. Below we list the proximity operators we use in our algorithm, and derive the one related to the smooth contour prior:

(a) Proximity operator of the data fitting term follows directly from the definition of proximity operators,

$$\begin{aligned} \text{prox}_{\mathbf{G}}^{\tau}(f_0) &= \arg \min_f \frac{\|f - f_0\|^2}{2\tau} + \mathbf{G}(f) \\ &= \arg \min_f \frac{\|f - f_0\|^2}{2\tau} + \|g_1 - M(\rho \otimes f)\|^2. \end{aligned} \quad (14)$$

This is a linear least-squares minimization problem, which we solve using the conjugate gradient method.

(b) Proximity operator of the total variation term is pointwise shrinkage [17],

$$\text{prox}_{\mathbf{F}_{\text{TV}}^*}^{\sigma}(y_0) = \frac{y_0}{\max(1, |y_0|)}. \quad (16)$$

(c) Proximity operator of the convex conjugate function  $\mathbf{F}_{\mathcal{E}}^*$  can be derived using Moreau’s Identity [17]. Moreau’s Identity relates the proximity operator of a convex conjugate function (e.g.  $\mathbf{F}_{\mathcal{E}}^*$ ) with the proximity operator of the original function (e.g.  $\mathbf{F}_{\mathcal{E}}$ ), and we get,  $\text{prox}_{\mathbf{F}_{\mathcal{E}}^*}^{\sigma}(\cdot)$  in terms of  $\text{prox}_{\mathbf{F}_{\mathcal{E}}}^{1/\sigma}(\cdot)$ ,

$$\text{prox}_{\mathbf{F}_{\mathcal{E}}^*}^{\sigma}(y_0) \equiv y_0 - \sigma \text{prox}_{\mathbf{F}_{\mathcal{E}}}^{1/\sigma} \left( \frac{y_0}{\sigma} \right), \quad (17)$$

where the proximity operator of the original function  $\text{prox}_{\mathbf{F}_{\mathcal{E}}}$  follows directly from the definition, and we get,

$$\begin{aligned} \text{prox}_{\mathbf{F}_{\mathcal{E}}^*}^{\sigma}(y_0) &\equiv y_0 - \sigma \left( \arg \min_y \frac{\sigma}{2} \left\| y - \frac{y_0}{\sigma} \right\|^2 + \frac{\lambda_{\mathcal{E}}}{2} \|y\|^2 \right) \\ &= \frac{\lambda_{\mathcal{E}}}{\sigma + \lambda_{\mathcal{E}}} y_0. \end{aligned} \quad (18)$$

This completes our algorithm. Below we discuss our prior.

## Modified smooth edge guided interpolation

In this section, we adapt the single-image super-resolution method [12] for our problem. [12] can be described as a sliding-window blur kernel estimation followed by a re-application of these kernels in the super-resolved grid. Our image data sampling strategies are different from theirs, and consequently we modify the kernel estimation process below.

We note that [12] uses a 4-neighborhood kernel. Over some sliding window  $\mathbf{W}$ , usually of size  $10 \times 10$  centered around each pixel, they estimate a *nontrivial* convolution kernel  $\mathbf{k}$  of size  $\approx 2 \times 2$  that  $\mathbf{W}$  is locally invariant to, i.e.,

$$\mathbf{W} \approx \mathbf{k} \otimes \mathbf{W}, \quad \mathbf{k} \neq \mathbf{I}. \quad (19)$$

This means that the kernel  $\mathbf{k}$  describes the local dominant direction of smoothness inside of the window  $\mathbf{W}$ . [12] then uses these estimated kernels for super-resolution. In their case, image pixel data is sampled evenly along both dimensions on a regular grid, and as a result a 2D disk-shaped support logically boils down to the 4-neighbor support (Fig. 6a)

For our problem, we use a 6-neighborhood instead. We observe that in the worst case when scene dynamic range is too wide, two very different exposure indices have to be used to capture as much of the scene information as possible. The two vastly different exposure settings will then capture two orthogonal sets of regions in the scene. In this case up to half of the pixel data would be lost due to clipping or noise (e.g., Fig. 1e). Let us denote the resulting vertical low-pass filter with  $\Gamma$ . Then the contents of a window with the same size in our case is  $\Gamma \otimes \mathbf{W}$ . Since convolution is commutative and  $\Gamma \otimes \Gamma = \Gamma$ , we get from (19),

$$\Gamma \otimes \mathbf{W} \approx (\Gamma \otimes \mathbf{k}) \otimes (\Gamma \otimes \mathbf{W}). \quad (20)$$

The resulting vertically low-passed kernel  $\Gamma \otimes \mathbf{k}$  has to have twice the size in the vertical dimension compared to  $\mathbf{k}$ , i.e., of size  $\approx 4 \times 2$ , and this yields a 6-neighborhood as demonstrated in Fig. 6b.

Note that this change of kernel size does not change the other parts of [12]. In particular, the order of magnitude speed-up proposed in [12] still applies to our method.

## Results

We show a few results in Figure 7. For each test case, we show two simulated exposures of our reconstructed HDR image.

The simulated short exposure images show details in the bright areas. In Fig. 7a-3 and Fig. 7b-3, our method has fully restored details inside of the bright regions a SDR camera would fail to capture. For comparison, Fig. 7a-2 shows the standard dynamic range image in full brightness while Fig. 7b-2 has its brightness matched with the simulated short exposure Fig. 7b-3.

The simulated long exposures demonstrate that the dark image areas have been restored relatively well. Note that we chose parameters such that we retain most of the noise in the dark areas. This is because strong denoising can potentially remove detail as well as noise. This is why the dark areas in simulated long exposures in Fig. 7a-4 and Fig. 7b-4 appear noisy.

In the examples presented, for the expmul capture of the input images we used dual ISO settings {100, 800} with an effective DR gain of 8x compared to a standard dynamic range image.



**Figure 7:** Results. The grey image on the left column shows the input exposure-multiplexed Bayer image. For our HDR reconstruction, both a short and a long simulated exposure is shown, for the demonstration of the quality of our reconstruction. Details are shown in the blown up insets.

## Conclusion

We have presented a robust method to reconstruct the latent high dynamic range image from a single exposure-multiplexed image. We use state-of-the-art global optimization techniques and a fast image prior inspired by recent single-image super-resolution research [12]. We have presented results produced by our algorithm and demonstrated the performance of our method.

## References

- [1] “Photography – Digital still cameras – Determination of exposure index, ISO speed ratings, standard output sensitivity, and recommended exposure index,” International Organization for Standardization, Geneva, CH, Standard, Apr. 2006. 1
- [2] P. Debevec and J. Malik, “Recovering high dynamic range images,” in *proceeding of the SPIE: Image Sensors*, vol. 3965, 1997, pp. 392–401. 1
- [3] T. Mitsunaga and S. K. Nayar, “Radiometric self calibration,” in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1. IEEE, 1999. 1
- [4] S. W. Hasinoff and K. N. Kutulakos, “A layer-based restoration framework for variable-aperture photography,” in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–8. 1
- [5] S. W. Hasinoff, D. Sharlet, R. Geiss, A. Adams, J. T. Barron, F. Kainz, J. Chen, and M. Levoy, “Burst photography for high dynamic range and low-light imaging on mobile cameras,” *ACM Transactions on Graphics (TOG)*, vol. 35, no. 6, p. 192, 2016. 1
- [6] J. Hu, O. Gallo, K. Pulli, and X. Sun, “Hdr deghosting: How to deal with saturation?” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1163–1170. 1
- [7] S. K. Nayar and T. Mitsunaga, “High dynamic range imaging: Spatially varying pixel exposures,” in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1. IEEE, 2000, pp. 472–479. 1
- [8] F. Heide, M. Steinberger, Y.-T. Tsai, M. Rouf, D. Pajak, D. Reddy, O. Gallo, J. Liu, W. Heidrich, K. Egiazarian, J. Kautz, and K. Pulli, “FlexISP: A flexible camera image processing framework,” *ACM Trans. on Graph.*, vol. 33, no. 6, p. 231, 2014. 1
- [9] a1ex. (2013) Dynamic range improvement for some Canon dSLRs by alternating ISO during sensor readout. [Online]. Available: [http://acoutts.com/a1ex/dual\\_iso.pdf](http://acoutts.com/a1ex/dual_iso.pdf) 1, 2
- [10] ——. (2013) Dual ISO - massive dynamic range improvement, magic lantern forum. [Online]. Available: <http://www.magiclantern.fm/forum/?topic=7139.0> 2
- [11] S. Hajisharif, J. Kronander, and J. Unger, “Adaptive dualiso hdr reconstruction,” *EURASIP Journal on Image and Video Processing*, vol. 2015, no. 1, p. 41, 2015. 2
- [12] M. Rouf, D. Reddy, K. Pulli, and R. Ward, “Fast edge-directed single-image super-resolution,” in *Proc. IS&T Intl. Symposium on Electronic Imaging 2016, in press*, 2016. 2, 4, 5, 6
- [13] A. Levin, R. Fergus, F. Durand, and W. Freeman, “Image and depth from a conventional camera with a coded aperture,” vol. 26, no. 3, p. 70, 2007. 2, 3
- [14] B. Olshausen and D. Field, “Emergence of simple-cell receptive field properties by learning a sparse code for natural images,” *Nature*, vol. 381, pp. 607–609, 1996. 2, 3
- [15] B. E. Bayer, “Color imaging array,” Jul. 20 1976, uS Patent 3,971,065. 3
- [16] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992. 4
- [17] A. Chambolle and T. Pock, “A first-order primal-dual algorithm for convex problems with applications to imaging,” *Math. Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011. 4, 5